DOCUMENT RESUME

ED 173 425 TM 009 513

AUTHOR Owston, Ronald D.
TITLE A Monte Carlo Comparison of Learning Hierarchy Validation Techniques.
PUB DATE Apr 79
NOTE 12p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Associative Learning; Complexity Level; Mathematical Models; Organization; Probability; Research Methodology; Simulation; *Skill Development; Task Analysis; Taxonomy; *Transfer of Training; *Validity
IDENTIFIERS *Learning Hierarchies; Maximum Likelihood Estimation; *Monte Carlo Methods

ABSTRACT

The development of a probabilistic model for validating Gange's learning hierarchies is described. Learning hierarchies are defined as paired networks of intellectual tasks arranged so that a substantial amount of positive transfer occurs from tasks in a lower position to connected ones in a higher position. This probabilistic validation technique is compared to two other selected validation techniques on 32 experimental populations representing valid and invalid hierarchies, various probabilities of measurement error, and several levels of skill difficulty. From each of the populations described, 1,000 random samples of four different sizes were generated by computer to obtain estimates of the probability of making an incorrect decision about the validity of the hierarchy for each technique. Results support the hypothesis that the developed technique is more accurate than the comparison techniques over all population-sample size combinations studied. (Author/CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED173425

Session No. 27.14

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

A MONTE CARLO COMPARISON OF LEARNING HIERARCHY VALIDATION TECHNIQUES

by

Ronald D. Owston

University of New Brunswick

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Ronald D. Owston

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, April 1979.

TM009 513

ABSTRACT

The development of a probabilistic learning hierarchy validation technique is described. The technique is compared to two selected validation techniques on 32 experimental populations representing valid and invalid hierarchies, various probabilities of measurement error and several levels of skill difficulty. From each of the populations, 1000 random samples of four different sizes were generated by computer to obtain estimates of the probability of making an incorrect decision about the validity of the hierarchy for each technique. Results supported the hypothesis that the developed technique is more accurate than the comparison techniques over all population-sample size combinations studied.

This report is based on the author's doctoral dissertation submitted to the University of Ottawa in 1978. The author wishes to acknowledge the assistance of Marvin Boss, Martin Cooper, and Jean-Paul Dionne.

Learning hierarchies are networks of tasks arranged in a manner such that a substantial amount of positive transfer occurs from tasks in a lower position in the network to connected ones in a higher position (Gagné, 1970). Two general methods have been used to validate postulated hierarchies. Gagné and White (1974) refer to them as the positive and negative methods. The positive method is based on the classic transfer model which implies that given possession of the subordinate skills, the learner is more likely to acquire the higher skill. The negative method rests on the view that the learner cannot acquire the higher skill unless all relevant subordinate skills are possessed. Although negatively validated hierarchies do not establish sufficient conditions for positive transfer, they do indicate the minimum set of skills the learner must possess to learn the higher skill. Thus most researchers have used the negative method as it is less time-consuming and requires fewer subjects than the positive method.

The negative validation method requires a group of subjects to be taught each skill in the hierarchy and then tested on possession of each skill immediately after (White, 1974a). The test data are then subjected to analysis by any one of a number of available validation techniques. Gagné and Paradise (1961) first proposed the proportion of positive transfer as a deterministic validation technique. This was followed by other deterministic techniques designed specifically for learning hierarchy validation, such as the adequacy, consistency and completeness ratios (Walbesser, 1968), the order ratio (Phillips, 1971), the conditional item difficulty index (Airaisian, 1971), ordering theory (Airaisian and Bart, 1973) and the difference ratio (Durell, 1973). In addition to these techniques, standard statistical procedures have been used to analyze hierarchy test data. These include the phi correlation coefficient, the chi-squared test of independence, and Gutman scalogram

analysis. Of all of these deterministic techniques, however, the Walbesser ratios appear to be the most frequently used.

White (1973, 1974b) called for a discontinuation of the use of deterministic techniques for one or more of the following reasons: (1) they are often severely affected by skill difficulty; (2) they do not take into account errors of measurement; (3) they do not have a known sampling distribution, and (4) the determination of the criterion for the validity of a hierarchy is often arbitrary. In place of these techniques, a probabilistic technique known as the test of inclusion (White and Clark, 1973) was proposed by White.

While the White and Clark test of inclusion eliminated many of the weaknesses of previous deterministic techniques, two shortcomings in this technique are evident. First, although errors of measurement are taken into account by the model, the parameter estimation procedure requires several undesirable assumptions to be made about the measurement error parameters. These assumptions are that the probability of an individual correctly guessing the answer to a question testing the subordinate skill is zero, and that there is no chance of an individual blundering on a question testing possession of the superordinate skill. The other is that the parameter estimation procedure is based only upon marginal totals of a 3x3 contingency table formed by using two questions to test possession of each of two skills. This procedure reduces the complexity of the parameter estimation, however, as a result valuable information is lost by ignoring the cell frequencies in the contingency table. These shortcomings imply the need for a theoretically more rigorous validation technique and suggest the manner in which one may be developed.

The Technique MAX

The technique MAX (Owston, 1978) is based on the original White and

Clark model. This model uses eight parameters, seven of which are independent. The parameters are P_0, P_I, P_{II} and P_B, the proportion of the population possessing neither skill in the hierarchy, only skill I, only skill II and both skill I and skill II; θ_a and θ_b are respectively the probability of an individual answering a skill I question correctly while truly possessing the skill and the probability of an individual correctly guessing the answer to a skill I question while not truly possessing the skill; and θ_c and θ_d are the skill II parameters corresponding to θ_a and θ_b.

Maximum likelihood estimates of the independent parameters in the White and Clark model are obtained by Fisher's method of scoring (Rao, 1968) directly from cell frequencies. This is an iterative procedure requiring rough approximations of the parameters at the beginning. In practice, estimates obtained by the White and Clark marginal totals method provide good initial starting values. From the estimate of the proportion of the population possessing only the superordinate skill ($\hat{P}_{II}$) and its variance, a confidence interval is then constructed. If the confidence interval includes zero, the conclusion is that there is no reason to assume that the hierarchy is not valid; otherwise, if the interval includes zero, the conclusion is that the hierarchy is not valid.

Theoretically, MAX appears superior to the White and Clark technique as no restrictions are placed on the value of the measurement error parameters (other than by definition they are between 0 and 1 inclusive), and parameter estimates are obtained from cell frequencies. Both of these techniques, however, appear superior to deterministic techniques.

Method

The relative accuracy of MAX, the White and Clark technique and the Walbesser ratios was investigated in this study using Monte Carlo methods. A

total of 32 experimental populations were used. These populations represented four levels of measurement error (5%, 10%, 20% and 25%) and four levels of skill difficulty (d_I = .95, .70, .40, .10; d_{II} = .90, .50, .25, .05) for pairs of skills that are hierarchically related (i.e. P_{II} = 0) and positively correlated but not hierarchical (ϕ = .25).

From each of the 32 experimental populations, 1000 random samples of size 25, 40, 100 and 400 were generated by computer. Each technique was applied to each sample to determine whether or not the hierarchy was valid. Two decision criteria were used with each technique: a 90% and 95% confidence interval for MAX; α = .05 and .10 for the White and Clark Technique; .85 for each of the completeness, consistency and adequacy ratios and .50 for the completeness ratio, .70 for the consistency ratio and .85 for the adequacy ratio of the Walbesser technique. The proportion of incorrect decisions (out of 1000) made with a given technique provided an estimate of the probability of making an incorrect decision about the validity of valid and invalid hierarchies. These probabilities were defined as Prob(I-V) and Prob(V-I) respectively.

In order to assess the relative accuracy of the various validation techniques criteria were established for Prob(I-V) and Prob(V-I). The most accurate technique was defined as the one with which simultaneous Prob(I-V) was at least .20 or less, and Prob(V-I) was at least .05 or less, most often over all population-sample size combinations studied. While these criteria are arbitrary, they do imply that to make an incorrect decision about an invalid hierarchy is approximately four times as serious as to make an incorrect decision about a valid hierarchy.

Results and Discussion

The obtained values of Prob(V-I) and Prob(I-V) are too numerous to list in this report but can be obtained from Owston (1978). Generally speaking,

Prob(I-V) approached one with the Walbesser ratios for easy skills and zero for difficult skills. The opposite occurred with Prob(V-I) with the Walbesser ratios: Prob(V-I) approached zero for easy skills and one for difficult skills. Neither the White and Clark technique nor MAX were affected by skill difficulty to the same extent. Both of these techniques were affected by sample size, however. As expected, samples of size 25 resulted in the least accuracy and samples of size 400, the greatest accuracy. With the White and Clark technique Prob(I-V) ranged from 0 to .3, Prob(V-I) from 0 to .96. Whereas with MAX, Prob(I-V) ranged from 0 to .08, Prob(V-I) from 0 to .37.

When validating a hierarchy, the researcher does not know the true state of validity of the hypothesized relationships between pairs of skills. Therefore, the risk of committing either kind of decision error with a given technique must be known. That is, for a given sample size and approximate level of skill difficulty and probability of measurement error, the researcher must have an estimate of Prob(V-I) and Prob(I-V) for the validation technique being used. This kind of information was obtained as estimates of Prob(V-I) and Prob(I-V) were found for the three validation techniques for valid and invalid hierarchies having the same sample sizes, levels of skill difficulty and probabilities of measurement error. These findings are summarized in Table I.

Shown in Table I is the number of times the criteria of Prob(I-V) $\leq$ .20 and Prob(V-I) $\leq$ .05 could be satisfied with each technique. As can be seen from this table the Prob(I-V) criterion was satisfied with MAX for all 64 population-sample size combinations for invalid hierarchies that were studied. The criterion was satisfied 60 and 51 times for the White and Clark technique for α = .05 and α = .10 respectively and not at all for the Walbesser ratios.

TABLE I

Number of Times Validation Techniques Satisfied the Decision Error Criteria on Valid Connections, Invalid Connections, and on Both Valid and Invalid Connections Simultaneously[a].

Technique	Number of Times Criteria Satisfied[b]		
	Valid Connections	Invalid Connections (ϕ= .25)	Valid and Invalid Connections Simultaneously
White and Clark (α = .05)	60	11	9
White and Clark (α = .10)	31	13	6
MAX (95% conf. interval)	64	12	12
MAX (90% conf. interval)	64	14	14
Walbesser ratios (.85)	0	48	0
Walbesser ratios (consist. = .70, comp. = .50, adeq. = .85)	0	48	0

[a]The decision error criteria were Prob(I-V) $\leq$ 0.20 for valid connections, Prob(V-I)$\leq$ 0.05 for invalid connections and both Prob(I-V) $\leq$ 0.20 and Prob(V-I)$\leq$ 0.05 for valid and invalid connections simultaneously.

[b]The maximum possible number of times the criteria could be satisfied was 64 in all three cases.

With invalid hierarchies, the Prob(V-I) criterion was satisfied most often using the White and Clark technique ($\alpha = .05$).

The data were next examined to determine how often both the Prob(I-V) and Prob(V-I) criteria were satisfied simultaneously. The criteria were simultaneously satisfied 14 and 12 times with MAX (with 90% and 95% confidence intervals respectively), 9 and 6 times with the White and Clark technique (with $\alpha = .05$ and $\alpha = .10$ respectively) and not at all for the Walbesser ratios. These findings suggest that MAX is generally superior to the White and Clark technique for validating hierarchies and that the Walbesser ratios are virtually of no use. In fact, a closer examination of the data showed, moreover, that on no occasion could the Prob(I-V) and Prob(V-I) criteria be simultaneously satisfied with the White and Clark technique and not with MAX.

In conclusion, this study supports the criticisms of deterministic validation made by White (1973, 1974b). Consequently every effort should be made to use probabilistic techniques, preferably MAX, whenever feasible. Nevertheless, caution still must be exercised when using probabilistic techniques as with small sample sizes and high probabilities of measurement error, decision errors may prove excessive.

REFERENCES

Airaisian, P.W. A method for validating sequential instructional hierarchies. Educational Technology, December 1971, 11, 54-56.

Airaisian, P.W. and Bart, W.M. Validating a priori instructional hierarchies. Journal of Educational Measurement, 1975, 3, 163-172.

Durell, A.B. A comparison of measures for validating learning hierarchies. Unpublished doctoral dissertation, University of Toronto, 1973.

Gagné, R.M. The conditions of learning. 2nd ed. New York: Holt, Rinehart and Winston, 1970.

Gagné, R.M. and Paradise, N.E. Abilities and learning sets in knowledge acquisition. Psychological Monographs, 1961, 75 (Whole No. 518).

Gagné, R.M. and White, R.T. Past and future research on learning hierarchies. Educational Psychologist, 1974, 11 , 19-28.

Owston, R.D. A comparison of learning hierarchy validation techniques using simulated data. Unpublished doctoral dissertation, University of Ottawa, 1978.

Phillips, E.R. Validating learning hierarchies for sequencing mathematical tasks. Unpublished doctoral dissertation, Purdue University, 1971.

Rao, C.R. Linear statistical inference and its applications. New York: Wiley, 1968.

Walbesser, H.H. An evaluation model and its application. American Association for the Advancement of Science: Washington, 1968.

White, R.T. Research into learning hierarchies. Review of Educational Research, 1973, 43, 361-374.

White, R.T. A model for validation of learning hierarchies. Journal of Research in Science Teaching, 1974, 11 , 1-3. (a)

White, R.T. Indexes used in testing the validity of learning hierarchies. Journal of Research in Science Teaching, 1974, 11 , 61-66. (b)

White, R.T. and Clark, R.M. A test of inclusion which allows for errors of measurement. Psychometrika, 1973, 38, 77-86.